Les probabilités sont-elles malades?

D'après un article paru dans « Le trésor de Tonton Lulu » - Editions Archimède

Congrès SBPMef 2016 – Michelle Solhosse

Le dépistage systématique d'une maladie sur base de prélèvement sanguin au sein d'une population coûte cher.

Confrontée à ce problème lors de la mobilisation des soldats pour la seconde guerre mondiale, l'US Army a comparé deux méthodes d'investigation :

- l'analyse individuelle : chaque prélèvement est analysé
- l'analyse collective : la population à étudier est répartie en groupes de ***n*** individus,

Dans ce cas, dans chaque groupe, les prélèvements sont mélangés et on analyse ce mélange.

Si le test est négatif

- aucun membre du groupe n’est malade
- une seule analyse

Si le test est positif

- au moins une personne du groupe est malade!
- ***n*** analyses supplémentaires

Comparer les deux méthodes

- La méthode collective est-elle rentable?
- A quelles conditions?
- Nombre de personnes dans chaque groupe ?

1. Simulation avec le tableur de TI Nspire

- avec un tableur : simulation de 100 groupes de ***n*** personnes (1< ***n*** <10).

- probabilité de n'être pas malade est ***p***.

p =.96

n =8.

1 = sain

0=malade

		B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	/gr'	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	'nb-in...	'gsain	'nb-gsain	
2	1.	0.	1.	1.	1.	1.	1.	1.	1.	_	_	7.	0.	66.	
3	2.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
4	3.	1.	0.	1.	1.	1.	1.	1.	1.	_	_	7.	0.	'nb-analyses'	
5	4.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.	372.	
6	5.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
7	6.	1.	1.	1.	1.	1.	0.	1.	1.	_	_	7.	0.	population testée...	
8	7.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.	800.	
9	8.	1.	1.	1.	1.	1.	1.	0.	1.	_	_	7.	0.		
10	9.	1.	0.	0.	1.	1.	1.	1.	1.	_	_	6.	0.	nb moyen d'analy...	
11	10.	1.	1.	1.	1.	1.	1.	1.	0.	_	_	7.	0.	0.465	
12	11.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
13	12.	1.	1.	0.	1.	1.	1.	1.	1.	_	_	7.	0.		
14	13.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
15	14.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
16	15.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
17	16.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
18	17.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
19	18.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
20	19.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
21	20.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		
22	21.	1.	1.	1.	1.	1.	1.	1.	1.	_	_	8.	1.		

N15

2. Point de vue théorique

- La probabilité pour que tous les individus d'un groupe de ***n*** personnes soient sains est p^n

  - <u>une seule analyse à effectuer</u>.

- La probabilité pour qu'une personne d'un tel groupe soit malade est $1 - p^n$

  - <u>***n***+1 analyses à effectuer</u>.

Le nombre d'analyses à effectuer pour un groupe de ***n*** personnes est une variable aléatoire ***X*** dont la loi de distribution est :

xi	***1***	***n+1***
pi	p^n	$1-p^n$

L'espérance mathématique est calculée par :

$$e = 1 \cdot p^n + (n+1) \cdot \left(1 - p^n\right) = 1 + n \cdot \left(1 - p^n\right)$$

La moyenne du nombre d'analyse à effectuer par personne est :

$$m = \frac{1}{n} + \left(1 - p^n\right)$$

Gain ou perte réalisé par individu :

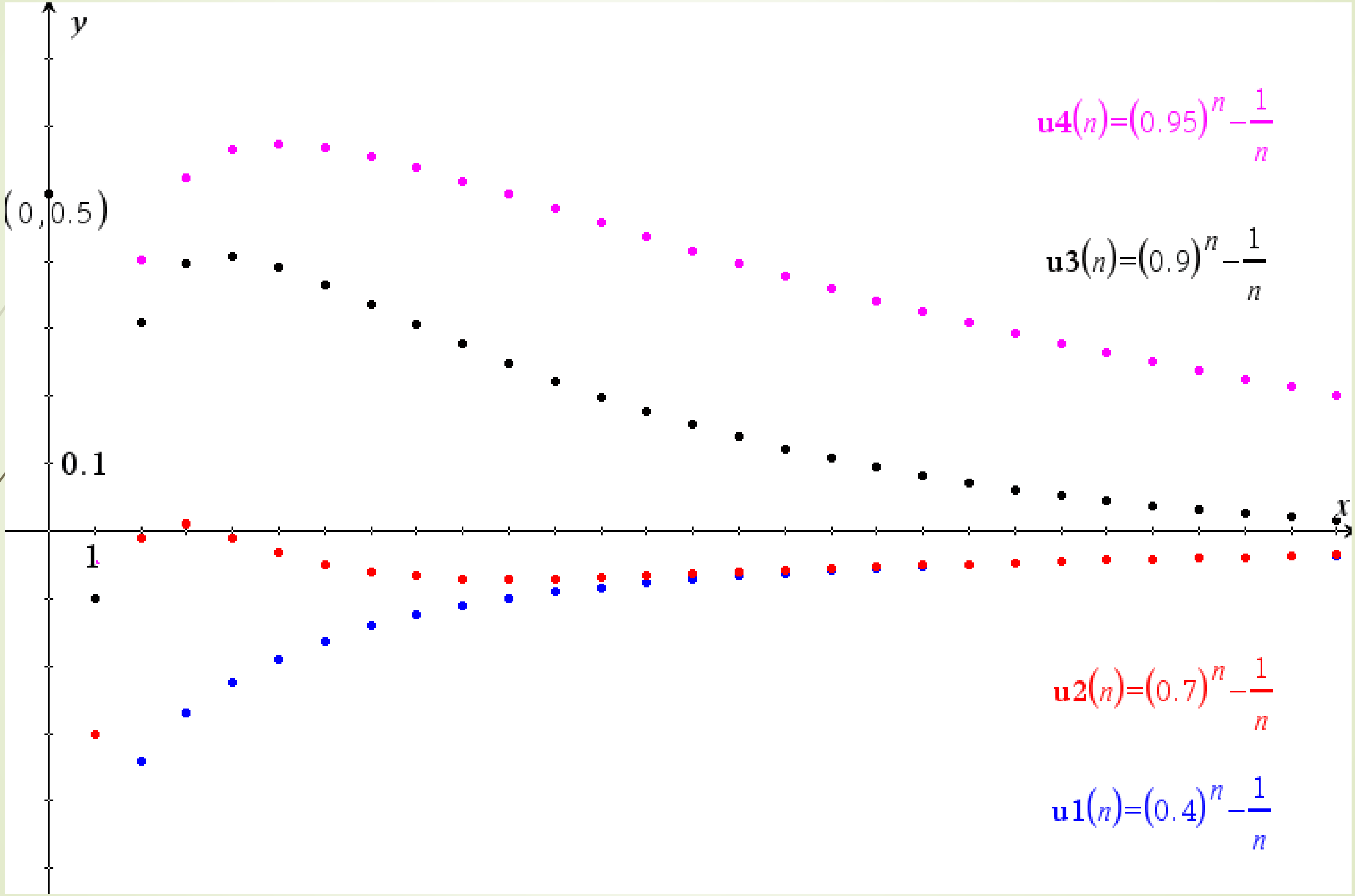
$$1 - m = p^n - \frac{1}{n}$$

3. Etude d’une suite

- L'expression qui nous intéresse est

$$u(n) = p^n - \frac{1}{n}$$

Observons ces suites pour différentes valeurs de ***p*** dans l’application graphique de TI Nspire.



y
(0, 0.5)
0.1
1
x
$u4(n)=(0.95)^n-\frac{1}{n}$
$u3(n)=(0.9)^n-\frac{1}{n}$
$u2(n)=(0.7)^n-\frac{1}{n}$
$u1(n)=(0.4)^n-\frac{1}{n}$

4. Caractéristiques de ces suites

- $u_1 < 0$

- $u_2 > u_1$ car $u_2 - u_1 = p^2 - \frac{1}{2} - p + 1 = p^2 - p + \frac{1}{2}$

  et ce trinôme est toujours positif

- $u_3 > u_2$ car $u_3 - u_2 = p^3 - \frac{1}{3} - p^2 + \frac{1}{2} = p^3 - p^2 + \frac{1}{6}$

  et ce polynôme n'admet pas de zéros dans R+

- à partir de u_4 ?

5. Conclusions

Grouper s'avère économique ? …

- on sait que $\boldsymbol{u}_1$ est toujours négatif, mais $\boldsymbol{u}_2$?

- $\boldsymbol{u}_2 > 0 \quad \Leftrightarrow \quad \boldsymbol{p}^2 - \frac{1}{2} > 0$ c'est-à-dire si $\boldsymbol{p} > \frac{\sqrt{2}}{2} \approx 0.705$

… mais , $\boldsymbol{u}_3 > \boldsymbol{u}_2$; dans ce cas, autant grouper par trois que par deux!

Autres conclusions

- Considérons donc le cas où $\boldsymbol{p} > \frac{\sqrt{2}}{2} \approx 0.705$

- Que peut-on dire de plus?

Par exemple : $\boldsymbol{u}_{10}$ est positif si $\boldsymbol{p}^{10} > \frac{1}{\sqrt[10]{10}} \approx 0.794$

Et $\boldsymbol{u}_{10}$ est négatif si $\boldsymbol{p}^{10} < 0.794$.

Par exemple, si $\boldsymbol{p} = 0.78$ ➡ maximum 9 termes de la suite positifs, parmi lesquels il faudra choisir le plus « économique ».

- Et plus généralement, $\boldsymbol{u_n} > 0$ dès que $\boldsymbol{p} > \dfrac{1}{\sqrt[n]{\boldsymbol{n}}}$
- $\sqrt[n]{\boldsymbol{n}}$ est une suite qui décroit vers 1 (pour ***n***>3);

  donc $\dfrac{1}{\sqrt[n]{\boldsymbol{n}}}$ croit vers 1, et quel que soit ***p***

  il y aura toujours une valeur de ***n*** telle que

  $\boldsymbol{p} < \dfrac{1}{\sqrt[n]{\boldsymbol{n}}}$
- Seuls un nombre fini de termes de la suite $\boldsymbol{u_n}$ sont positifs. C'est parmi ceux-ci qu'il faut choisir le maximal.